

# Introduction to Python on ASU RC HPC Systems

Jason Yalim

August 26, 2019



python<sup>TM</sup>

# **Why High Performance Computing (HPC)**

Pool of capital investment into state of the art hardware and software

Team(s) of system administrators and IT to maintain hardware and software

Advantages of remote computation

# The Statistics Student

Suppose statistics student requires cross validation study

1000 runs required

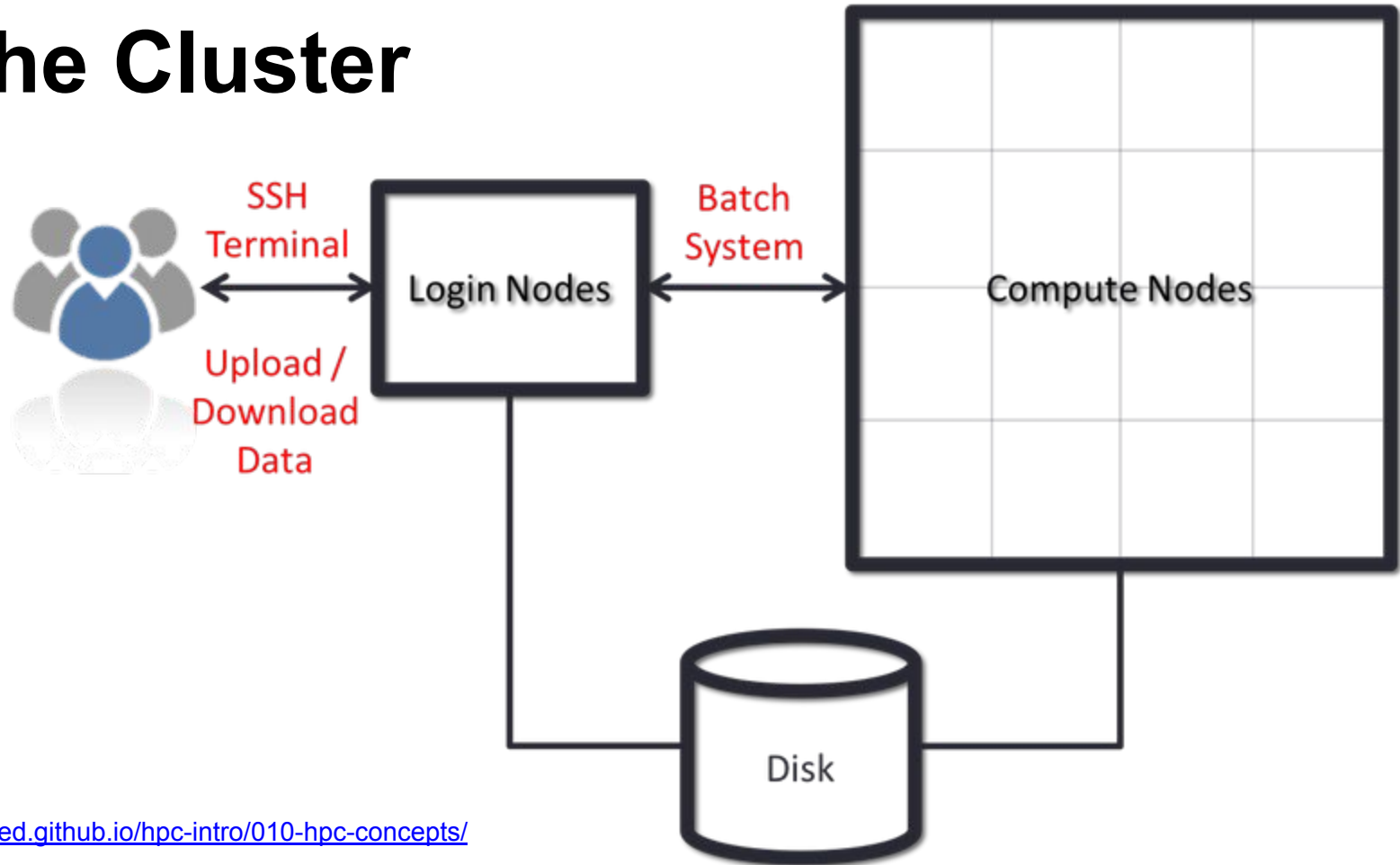
Each run takes 1 hour on laptop

41 days and 16 hours of compute time!

# Considerations for HPC

- **Speed.** With many more processing cores, often with higher performance specs, than a typical laptop or desktop, HPC systems can offer significant speed up.
- **Volume.** Many HPC systems have both the processing memory (RAM) and disk storage to handle very large amounts of data.
- **Efficiency.** Many HPC systems operate a pool of resources that are drawn on by a many users. In most cases when the pool is large and diverse enough the resources on the system are used almost constantly.
- **Cost.** Bulk purchasing and government funding mean that the cost to the research community for using these systems is significantly less than it would be otherwise.
- **Convenience.** Maybe your calculations just take a long time to run or are otherwise inconvenient to run on your personal computer. There's no need to tie up your own computer for hours when you can use someone else's instead.

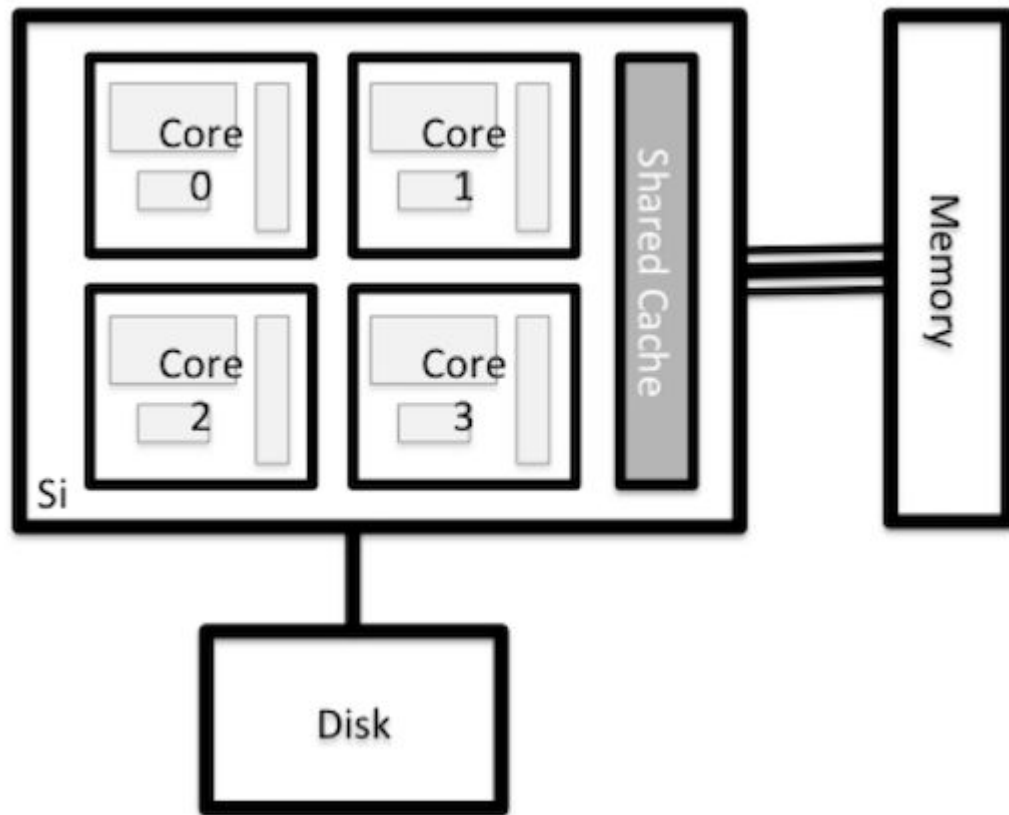
# The Cluster



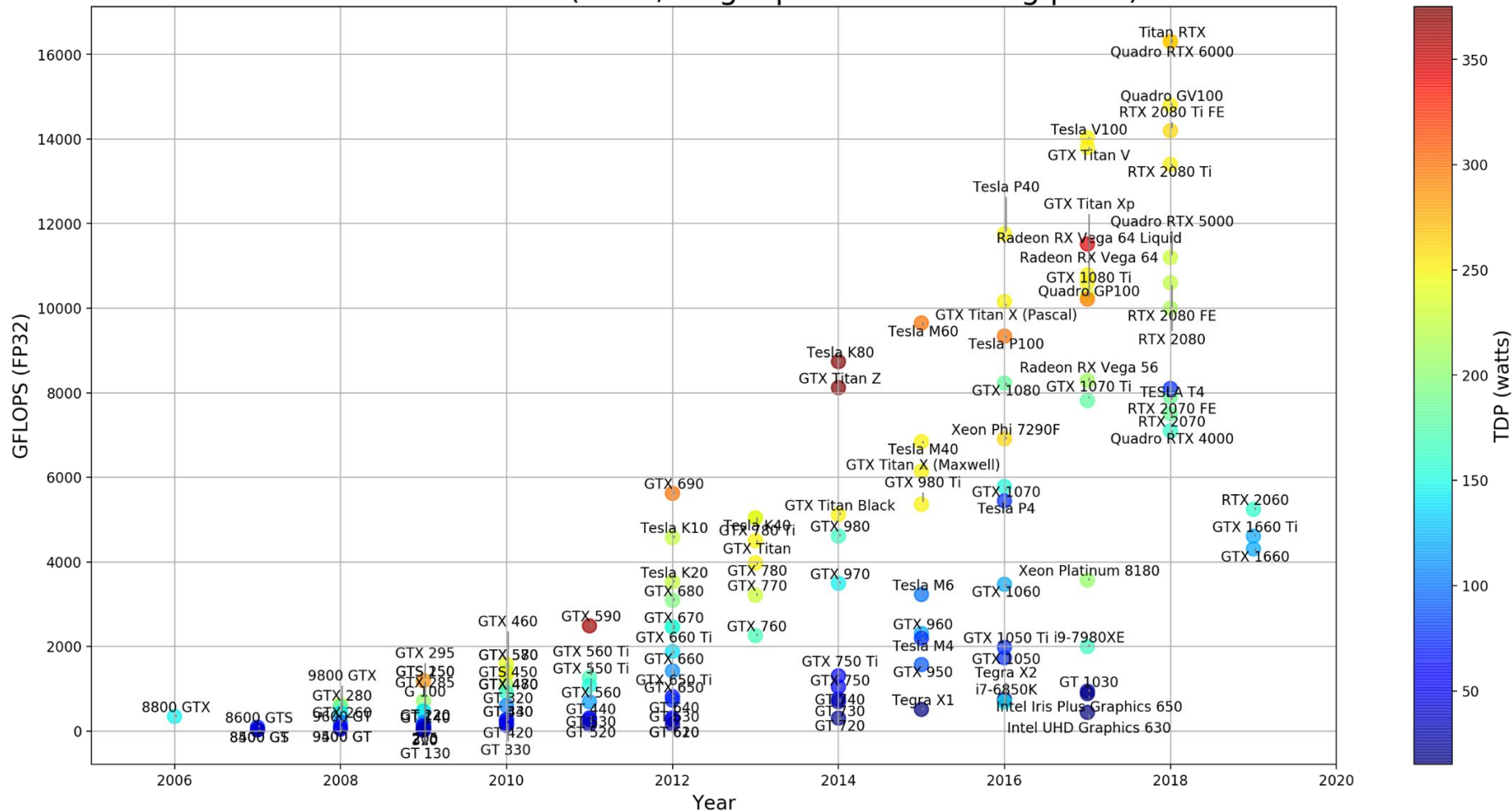
# Compute Nodes

[Agave](#) has nearly 350 compute nodes

Includes nearly 200 GPUs



# GPU Performance (FP32, single precision floating point)



# Upcoming Workshops (1/2)

Intro to ASU's HPC Cluster

September 9th

Biodesign Auditorium, 11 AM



Scan me



# Upcoming Workshops (2/2)

Data Handling with Python's  
numpy and Pandas on the ASU  
Research Computing Cluster

September 13th

Biodesign AL 10-14

2-3 PM

3-4 PM



Scan me

# Conclusion

For any assistance please contact: [support@hpchelp.asu.edu](mailto:support@hpchelp.asu.edu)

Office Hours: GWC546 1-4 PM Tuesdays and Wednesdays

Info: [researchcomputing.asu.edu](http://researchcomputing.asu.edu) [rcstatus.asu.edu/agave](http://rcstatus.asu.edu/agave)

**Supplemental Slides Follow**

# The Scheduler



1. Transfer input datasets to the HPC system (via the login nodes)
2. Create a job submission script to perform your computation (on the login nodes)
3. Submit your job submission script to the scheduler (on the login nodes)
4. Scheduler runs your computation (on the compute nodes)
5. Analyse results from your computation (on the login or compute nodes, or transfer data for analysis elsewhere)

# Data Plague

Suppose a genomics researcher is sequencing small datasets, 1000 each 500 MB

But new datasets are in, 20x as large (10 GB), 2x as many

500 MB file is already a burden on most machines!

# Idle Resources

Suppose a student with 4 core laptop is running a 2D PDE solver in serial

Needs to go to 3D and parallel solver can be implemented

Convert laptop into quasi-permanent solver/heater?

<https://epcced.github.io/hpc-intro/010-hpc-concepts/>

- **High Performance Computing (HPC).** Computing resources that allow people to solve their problems faster or treat larger problems than they would be able to using standard computing resources (e.g. a laptop, desktop or workstation). Usually implies some sort of *parallel computing*.
- **Parallel Computing.** The use of computing resources in parallel to speed up computation or treat larger computational problems.
- **Supercomputer.** Typically used to describe a very large HPC resource such as those found on the [Top500](#) list. Often uses the same technology as *compute clusters* but at a larger scale.
- **Compute Cluster.** Typically used to describe a smaller HPC resource than those referred to as *supercomputers*. Usually use exactly the same technology as supercomputers but on a smaller scale.
- **High Throughput Computing (HTC).** A subset of parallel computing where computing resources are used in parallel on many independent sub-tasks to increase the rate at which computation can be performed. For example, varying an input parameter (or input data) to a computation and running many copies simultaneously.
- **Cloud Computing.** Using remote computing resources *on demand*. Often associated with using public cloud computing resources provided by large internet corporations.